Tackling the Challenges of Artificial Intelligence for Public Communication:

Governance and Accountability Discourses About the Responsible Use of AI

Abstract

Artificial intelligence (AI) is rapidly transforming the landscape of public communication. However, while AI technologies allow for more participatory and dialogic processes, they can also facilitate destructive communication patterns including hate speech, online harassment, and dis- or misinformation. Such trends raise concerns about how to regulate the responsible use of AI in public communication. This contribution intends to reflect the challenges AI presents to public communication and examine the governance and accountability frameworks that seek to ensure transparency, fairness, and responsibility. To this end, the authors established a database of 435 national and international codes of ethics and other normative guidelines. These documents were subjected to an inductive thematic analysis to identify key themes and analyze differences and similarities from a comparative perspective. The empirical study reveals significant gaps in most of the documents but also collects suggestions on the way towards a coherent governance framework for AI in public communication.

Keywords

Artificial Intelligence, Hybrid Media System, Governance, Accountability, Public Communication, Public Sphere, Communication Ethics, Media, Journalism

Introduction

The current hybrid media ecosystem (Chadwick, 2017) has led to the emergence of numerous actors in public communication that shape, along with established news media, the information flows in our democracies. In this new digital environment, artificial intelligence (AI) is rapidly transforming the landscape of public communication, reshaping how information is created, disseminated, and consumed. Especially the advent of generative AI represents "a watershed moment across a range of professional fields, including and perhaps especially in communication and media industries" (Guzman and Lewis, 2024: 348). However, this technological shift raises concerns about AI's potentially dysfunctional impacts on public communication and the public arena (Jungherr and Schroeder, 2023), such as reinforcing harmful patterns like disinformation, incivility, and echo chambers. Research has explored AI's implications for both individual actors (Guzman and Lewis, 2020; Natale, 2021; Verdegem, 2021) and institutions (Coeckelbergh, 2024; de-Lima-Santos and Ceron, 2021;

Diakopoulos, 2019; Simon, 2022, 2024a), particularly regarding its effects on the public sphere and democratic processes (Jungherr, 2023).

A key challenge concerns defining AI. There is no universally accepted definition of AI, and current debates are strongly influenced by generative AI, even if AI systems have been used for quite some time (Simon, 2024a). AI is generally understood as simulating human intelligence and decision-making carried out by computer systems (Gunkel, 2020). For instance, the Steering Committee on Media and Information Society (CDMSI) of the Council of Europe refers to AI systems as "algorithmic system or a combination of such systems that uses computational methods derived from statistics or other mathematical techniques and that generates text, sound, image or other content or either assists or replaces human decision-making" (Council of Europe 2023: 7). According to the CDMSI, these systems do not constitute a single technology but rather a set of different, often interconnected tools designed to automate specific tasks, and as such, include also algorithmic forms. In addition, beyond a purely technical definition of AI, one should also take into account "that individuals and social actors understand AI in different ways and attribute varying potentials to the technology" (Vogler et al., 2025: 9). Perceptions of AI are therefore also socially constructed and play a key role in shaping which opportunities and risks are associated with it.

The continuous scrutiny of the technology's wider implications, for instance through algorithmic accountability journalism (Diakopoulos and Koliska, 2017), has not only raised awareness about the technology's societal impact and how media communicate with people (Jungherr and Schroeder, 2023; Simon, 2024b), but it has also influenced growing policy debates in the news media, in politics, and with civil society actors about how to regulate and ensure a responsible use of AI in relation to public communication. This article adds to the ongoing debate by exploring the role of AI in shaping public communication, focusing specifically on the way European media as well as civil society actors seek to ensure a responsible use of AI-driven tools in public communication through governance and accountability frameworks. Using a hybrid media ecosystem-perspective, we adopt a more expansive approach beyond news media emphasizing our concern with the pervasiveness of AI throughout a wide range of professional fields and communication industries that goes beyond news and journalism.

Grounded in governance theories (Dafoe, 2018) and media accountability frameworks (Eberwein et al., 2019), our study engages with broader debates on democracy and public communication, focusing on the ethical and responsible deployment as well as the development of AI systems. It presents findings from the EU-funded *ANONYMIZED PROJECT NAME*, which investigates AI governance mechanisms and their impact on public communication across Europe. The analysis draws on a database of 435 national and international codes of ethics and similar guidelines from media as well members of the civil society, as well as political and economic actors across eight European countries (Austria, Estonia, Finland, Hungary, the Netherlands, Lithuania, Slovenia, and Switzerland), detailing

governance and accountability strategies for fostering responsible public communication amid disruptive technological change. The article pursues three key objectives: (1) identifying the main themes in terms of the challenges AI presents to public communication and democratic dialogue as raised in the documents; (2) examining the mechanisms of AI governance; and (3) analyzing to what extent the documents diverge between the national and supranational level regarding the mentioned challenges and mechanisms.

Literature review

Chadwick's (2017) concept of the hybrid media ecosystem outlines a media environment where traditional journalism, digital platforms, and civil society intersect, creating a space for complex interactions that reshape public communication. Central to this concept is the role of AI and algorithms, which fundamentally alter how information is produced and how it circulates in the public sphere. Algorithms determine the visibility and prominence of news stories across platforms, often prioritizing content based on engagement metrics, with implications for what Chadwick terms the "information flows" within the hybrid ecosystem, as they can amplify certain narratives over others based on their appeal – or ideological stance – rather than their societal or journalistic relevance. This shift in gatekeeping reshapes the dynamics of public communication by privileging content that aligns with platform-specific engagement criteria, whether economically or politically-driven, potentially sidelining less sensational but socially important news.

Impact of AI on public communication and democracy

Jungherr and Schroeder (2023) argue that AI can alter public communication by shaping the visibility of societal issues and enabling spaces for both public and counterpublic formation. Central to their argument is the assertion that AI strengthens intermediary control over the public arena, as algorithmic content curation limits the scope and nature of discourse by prioritizing narratives over others, affecting in turn the transparency and inclusiveness of democratic discourse. As users increasingly consume news through platforms like social media, news aggregators, and search engines, the algorithms that curate (and, like Google's AI Overviews, to some extent produce) this content amplify stories that elicit high engagement. Consequently, AI-driven platforms not only reshape the information environment and user behavior (Jungherr and Schroeder, 2023) but also the social structures within the public sphere, as news consumption becomes increasingly fragmented along ideological lines, posing challenges for maintaining a balanced political discourse.

This environment also fosters what Chadwick describes as a "cross-pollination" of information, where user-generated narratives – or content produced by alternative media – can gain traction and influence public discourse, often without verification or fact-checking processes. Even more so after the recent announcement of Meta to discontinue third-party

fact-checking and to lift restrictions on topics previously deemed harmful (Kaplan, 2025). Consequently, Jungherr and Schroeder (2023) emphasize the need for transparency and accountability of AI systems within the public arena, stressing the importance of policy interventions that uphold democratic principles. Otherwise, if algorithms are not transparently managed and held to account, they will erode trust (Jungherr, 2023).

Similar discourses are brought forward in news and journalism. According to Simon (2024a), AI technologies are not only retooling journalistic practices, but by optimizing news production processes they redefine what journalism is and how it functions within society and democracy. For instance, news recommenders are increasingly used in the distribution of news and information (Mitova, 2023): as AI selects and prioritizes stories likely to drive engagement or foster ideological interests of owners and algorithm creators, it risks creating an information landscape dominated by sensational or divisive content, potentially distorting public perception and dialogue (idem). In fact, Helberger et al. (2020) show that AI's role in curating news content on social media platforms and in traditional news media raises significant concerns regarding public opinion formation. These concerns are reflected by the audience's critical stance towards the use of AI in journalism: a recent study in Switzerland (Vogler et al., 2023) showed that AI in news production is seen extremely critical by the audience. AI-driven processes often lack transparency, as algorithms operate based on complex data inputs that are often inaccessible or incomprehensible to the public, also known as the Black Box Problem in AI (Zednik, 2021). Both Simon (2024a) and (ANONYMIZED, 2024) stress the need for greater accountability and transparency in AI-driven journalism, warning that algorithmic bias may distort democratic discourse by amplifying some voices while marginalizing others.

AI, governance and accountability

The discussions over responsible development and use of AI systems are not unique to the hybrid media ecosystem but are essential in overall discussions about regulating AI. On a European level, there are three main guiding documents for AI ethics (European Commission, 2019; Organization for Economic Cooperation and Development OECD, 2019 and United Nations Educational, Scientific and Cultural Organization UNESCO, 2021) and with the AI act (European Parliament and Council, 2024) a specific law. All of them highlight relevant principles for AI systems, for example transparency, safety, privacy, and accountability.

Although the abovementioned documents seek to provide universal norms for AI governance in different domains, they rarely take into account the news media industry (Porlezza, 2024). In addition, it is also a challenge to interpret and apply important but abstract principles (see e.g. Jobin et al., 2019) to journalism, or even to establish a normative agreement on key values and principles (Sutrop, 2020).

A major concern within the hybrid media ecosystem is therefore the question of how to regulate AI responsibly and ensure accountability – often seen as one of the cornerstones of

the governance of AI (Novelli et al., 2023). Eberwein et al. (2019) argue that media accountability frameworks must evolve to address the unique challenges posed by AI, advocating for mechanisms that hold AI systems to specific standards such as transparency, fairness, and human oversight. Along the same argumentative line, Dafoe (2018) underscores the importance of developing regulatory structures that ensure responsible AI use by strengthening transparency, accountability, and public oversight as central principles. Frameworks that support accountability and oversight in public communication are essential as AI-driven decisions can significantly impact the public sphere and democratic processes. At the international level, first governance and policy initiatives dedicated specifically to generative AI have been launched, but the approach remains narrow and technocratic in what Ulnicane (2025) defines "governance fix".

To address the issue of adequate (self-)regulatory framework for accountable use of AI in public communication, this paper stems from an updated notion of accountability, understood as a relational and distributed process embedded in hybrid media ecosystems, where responsibility is shared across interconnected actors rather than assigned through linear chains of control. However, "a new understanding of accountability requires an awareness of the relational context within which responsibilities and duties develop. It demands an acknowledgement of the dynamic network of interactive relationships within which individuals and organizations are embedded in (their) environment, as well as a willingness to seriously consider the very consequential role and effect of expectations and perceptions within such a context" (Painter, 2007: 526). To ensure their effectiveness, accountability systems are typically employed and conceptualized in a manner that integrates performance standards, external monitoring, and the evaluation of outcomes, coupled with consequential rewards and sanctions (Sahlberg, 2010: 51). Such accountability systems can indirectly assist stakeholders (such as news media and journalists) in striving for optimal results (Sahlberg, 2010: 51) and in adhering to the mandates set forth by professional or legal codes (Painter, 2007: 517).

The literature on accountability underscores its dynamic and multifaceted nature, emphasizing the interplay between internal mechanisms – such as self-regulation, critical reflection, and community-driven processes – and external systems, including regulatory frameworks, monitoring tools, and public oversight (Edwards and Hulme, 1996; Hargreaves, 2008; Sahlberg, 2010; Kaldor, 2003). Effective accountability often involves building mutual trust and collaboration among stakeholders while fostering transparency and inclusivity, in particular in a hybrid news media ecosystem, where numerous actors contribute to the information flows in public communication. This "intelligent accountability" approach highlights the importance of relational and institutional mechanisms (Sahlberg, 2007). Intelligent accountability stresses therefore the principle of mutual responsibility (Sahlberg, 2010: 53–54). Abelmann and Elmore (1999) created a similar model of internal accountability which assumes that stakeholders embed accountability in their daily operations. Such an

“intelligent accountability” approach can ensure that stakeholders work effectively and efficiently toward both the public good and the development of their work.

Implementing these accountability frameworks, however, faces significant challenges. The absence of enabling environments – comprising supportive policies, legal systems, and (self-) regulatory structures – can hinder the development and institutionalization of effective accountability practices (Ahmad, 2009). Additionally, conflicts between internal and external accountability systems frequently emerge, particularly when organizational management practices or political objectives clash with professional ethical principles and societal expectations (Kaldor, 2003). In the realm of AI, these challenges become even more pronounced. While ethical guidelines and codes of conduct are promoted as means to ensure transparency and accountability in the development and use of AI (Corrêa et al., 2023; Becker et al., 2025), their implementation is often undermined by the profit-driven motives of corporate stakeholders (Steinhoff, 2023). Without robust governance frameworks that balance economic interests with societal needs, ethical commitments risk being reduced to AI ethics washing (Schultz and Seele, 2023; Wagner, 2018). Moreover, the complexity of AI systems, their rapid evolution, and their global and transnational reach further complicate the task of developing effective accountability mechanisms, not only in Europe, but at a global level.

Scholars emphasize therefore the necessity of fostering multi-stakeholder governance systems that engage diverse actors, including governments, legal institutions, civil society, and corporations (*ANONYMIZED*, 2023; van Es and Nguyen, 2024). This approach calls for a nuanced understanding of accountability as a social construct embedded in relational and institutional contexts (Painter, 2007). Such frameworks must not only regulate AI development but also create environments where accountability is shared, adaptive, resilient, and interactive (Rozgonyi, 2023). This involves addressing structural and systemic challenges, ensuring inclusivity in governance processes, and recognizing the critical role of political action alongside institutional partnerships (Ahmad, 2009). All this should be aimed at developing an AI governance framework that can lead to accountable artificial intelligence (AAI) in relation to public communication in democracies. By aligning these efforts with democratic principles, AAI can support responsible public communication and the safeguarding of public trust in the face of disruptive technological change.

The article aims to contribute to broader debates on democracy and public communication by exploring the ethical and responsible deployment of AI systems as defined by the Council of Europe in its Guidelines on the responsible implementation of artificial intelligence (AI) systems in journalism (2023). Specifically, it investigates the governance and accountability mechanisms – such as codes of ethics developed by different stakeholders such as media organizations, members of civil society as well as political and economic actors – that shape AI’s impact on public communication across Europe. The article seeks to identify the ethical issues associated with responsible use of AI, describe the relevant accountability or

governance mechanisms, and examine the challenges associated with their implementation. The article addresses the following research questions:

RQ1) Which AI-related challenges are covered in the documents collected?
RQ2) Which governance and accountability mechanisms emerge in the analyzed AI-related documents?
RQ3) How do the documents issued by national and supranational stakeholders differ?

Methodological approach

In order to answer the research questions, the authors conducted a comparative study of codes of ethics and other normative guidelines for public communication in eight European countries as part of the previously mentioned *ANONYMIZED* project: Austria, Estonia, Finland, Hungary, the Netherlands, Lithuania, Slovenia, and Switzerland. The selection of countries represents different types of media systems in Eastern and Western Europe (Dobek-Ostrowska et al., 2010; Hallin and Mancini, 2012; Peruško et al., 2020), and thus corresponds to the logic of a Most Different Systems Design – a common approach in comparative social science aimed at deriving universally valid insights from the broadest possible variety of cases (Landman, 2008). The heterogeneity of the countries included is intended to enhance the external validity of the study and, in turn, to increase the likelihood of generating findings that are transferable to other media systems. This strategy reflects the recognition that AI-related challenges are currently present across a wide range of national contexts around the globe – and that the search for appropriate regulatory responses must not, of course, be confined to individual media systems.

At the same time, the project deliberately focuses on societies with comparatively small populations, acknowledging that 'smallness' affects a country's cultural and linguistic diversity, the size and structure of its media market, as well as content distribution and consumption patterns (Kõuts-Klemm et al., 2024). *ANONYMIZED PROJECT NAME* follows the assumption that small countries are particularly suitable for investigating ethical standards and policies (including but not limited to those related to AI) because even local communication crises appear more pronounced and attract proportionately more attention than in large societies (*ANONYMIZED*, 2022). The problems of society's resilience when facing communication and information disorders become visible quicker and clearer and are thus easier to identify. On the other hand, small countries can react more flexibly to policy challenges and information disorders than big countries. From an empirical perspective – and of particular relevance for a multi-country research consortium – it should be noted that reaching relevant target groups is easier in smaller linguistic communities. In addition, fewer resources are required to identify pertinent samples of journalists, experts, educators, and others for research purposes, as they are typically well organized (*ANONYMZED*, 2022). All of this seems highly relevant for a systematic investigation of new types of governance challenges such as the question of responsible use of AI.

For the study presented below, a two-stage empirical research process was realized in all eight *ANONYMIZED PROJECT NAME* countries, combining quantitative and qualitative approaches:

Drawing on the media governance and accountability concepts outlined above, the *first stage* of the study involved systematic, theory-based data collection, indexing, and annotation in line with the logic of a document analysis (Prior, 2003). In this process, all research partners involved in the project identified examples of ethical codes and other guidelines for public communication in their own countries and at supranational level. To this end, various types of codes/guidelines were defined in advance, each of which was to be represented by at least two to three typical documents in each country (in detail: codes/guidelines for journalists, for advertisers, for PR professionals, for corporate communicators, for public institutions, for small-scale or individual media, for media users). Within these segments, different levels of codes/guidelines (macro, meso, micro) were also represented where possible, resulting in a heterogeneous mixture of documents with varying forms of institutional affiliation. Documents were only taken into account if they complied with the principles of self-regulation and accountability in public communication (Fengler et al., 2022). Legal texts in the sense of 'hard law' (i.e., binding legal instruments such as constitutions, regulations, and court rulings with enforceable authority) were therefore excluded from the data collection, as were numerous guidelines that primarily deal with interpersonal communication practices, since these appeared to be of only marginal relevance to the research interest of the *ANONYMIZED PROJECT NAME* study. Research partners collected a total of 435 national and international documents in this process, and most of them were subsequently made publicly available in an interactive database on the *ANONYMIZED PROJECT NAME* website[1] – unless their originators insisted on confidentiality. Due to the purposive selection strategy, the database cannot fulfill the claim of representativeness, but it can offer a broad and systematic overview of codes of ethics and additional guidelines for all types of public communication that does not exist in this form in previous research (see above). It goes without saying that such documents cannot capture the actual editorial practices in contexts of public communication, but only an ideal state to be aspired to. This well-known tension between ought and is, not uncommon in ethical discourses, will be revisited after the presentation of our findings – along with the question of the extent to which codes of ethics can ultimately serve as effective instruments for enforcing 'good communication conduct' (see also Fengler et al., 2015).

The data collection process was supplemented by standardized indexing and annotation, for which the collaborating researchers extracted relevant basic information and additional contextual data for all documents (including name of code, originator of code, country, original language, year of creation, year of last adaptation, length, type of code, main frame of accountability – as well as a binary coding used to indicate the relevance of the document for issues related to AI). All documents for which the latter category attested a reference to the theme of AI were considered for an in-depth analysis in the study summarized in this

paper. To validate the indexing by the national research partners, the entire *ANONYMIZED PROJECT NAME* database was scanned again using various search strings related to AI.[2]

After adjusting the preselection, a subsample of 63 documents with a specific bearing for governance and accountability discourses about AI remained (see Table 1). This subsample (corresponding to 14,5% of the total sample) forms the basis for the analysis presented below. Almost half (46,0%) of the AI-related codes and guidelines originate from supranational organizations; relevant documents at the country-level remain in the minority. It is therefore not surprising that a majority of 39 documents (61,9%) are written in English.[3] The most recent document are the Artificial Intelligence Guidelines by the Swiss Press Council, published in 2024; the document with the longest tradition is ICC Advertising and Marketing Communications Code by the International Chamber of Commerce (ICC), originally dating from 1937, which was last adapted in 2018/2024[4] to take account of new ethical issues in the context of AI.

Table 1: Number of AI-related documents per country and addressees

	Policy stakeholders	Journalism	Advertising and PR	Public institutions	Corporate	Media user	Other	Total
Austria	0	1	2	2	0	2	0	7
Estonia	0	0	0	0	0	0	0	0
Finland	0	3	0	1	0	0	0	4
Lithuania	0	3	0	1	0	0	0	4
Hungary	0	1	0	0	0	0	0	1
The Netherlands	0	5	0	1	0	0	0	6
Slovenia	0	0	0	0	0	0	0	0
Switzerland	0	8	0	1	3	0	0	12
International	6	2	3	4	0	13	1	29
Total	**6**	**23**	**5**	**10**	**3**	**15**	**1**	**63**

(source: own data)

While the first analytical steps followed a theory-based deductive procedure, the *second stage* of the empirical study was essentially organized as an inductive thematic analysis (Braun and Clarke, 2012). To this end, the entire subsample of 63 AI-related documents was subjected to a multi-stage coding process using the MAXQDA software (VERBI Software, 2021). Contributing coders highlighted: (1) whether AI is at the core of the documents' purpose – or whether it is only a peripheral aspect (i.e. one theme among others); (2) what are the predominant themes and typical areas of risks related to AI in the documents examined; (3) how the identified themes and risks can be merged into larger units of meaning; (4) which tools/mechanisms the documents propose to tackle the previously identified risks related to AI; and (5) what are potential good practices for media governance and accountability. The coding process was conducted in two separate subgroups within the collaborating research teams. For quality assurance purposes, 10% of the documents were

coded multiple times to ensure a consistent approach. The procedure did not reveal any notable ambiguities in the coding instructions, thereby indicating a high level of reliability in our analytical strategy. For the evaluation of the collected data, some of the previously coded basic categories (e.g., type of code; main frame of accountability) were also considered (see above). The distinction between national and supranational documents enabled a comparative evaluation of the corpus.

Results

Our examination of codes of ethics and other guidelines for public communication in eight European countries reveals that the relevance of responsible AI use is hardly reflected in these documents. Out of the 435 documents collected for the ANONYMIZED project, only 63 contain any references to AI and automation. Within this subsample, roughly one-third of the documents pay attention to AI-related issues in detail: while AI is a central focus in 20 documents, the remaining 43 examples make only peripheral references – and approximately one half of them contains no more than one relevant text passage. Some of the analyzed countries (Estonia and Slovenia) showcase no examples of AI-related codes at all, while in the other national samples at least a few central documents from tech-savvy media companies can be identified as pioneers in this field (e.g., the *Guideline for Artificial Intelligence* by the Austrian Press Agency APA or the *Declaration of Intent for Responsible Use of Artificial Intelligence in the Media* by the Dutch practitioner initiative Media Perspectives). The biggest cluster of documents with an in-depth reflection of ethical issues related to digital spheres and AI was issued by European or international organizations (such as the European Commission, the United Nations, and UNESCO), with some of them offering examples of soft-law and policy documents constituting practices of co-regulation rather than media accountability in a narrow sense.

AI-related challenges

Given the multifaceted nature of AI, there is no universally applicable approach to guide its entire lifecycle – consequently, the pertinent issues and risks vary depending on the specific area of application. Our inductive thematic analysis suggests three main categories of challenges related to AI, namely (a) those documents that arise in the development and operation of AI systems, (b) those that offer a legitimization for the use of AI, and (c) those related to the deployment of AI.

According to the analysis, the ethical reflection of AI systems is of utmost importance during the *development and operation stage*. Operators and owners of AI systems are already confronted with risks relating to the need for technical robustness of AI systems, inter alia potential for errors (‘hallucinations’ or ‘biases’), discrimination and the impact on data and privacy. In addition, it is necessary to consider international human rights law and standards, as well as the potential risks associated with overreliance on the system’s capabilities and

limitations by users. Microsoft Bing's *Approach to Responsible AI* addresses several additional concerns, including the possibility of jailbreaks, the creation of ungrounded AI-generated content, and the issue of how to address child users. These findings correspond to the European Commission's *Ethics Guidelines for Trustworthy AI*, which underscore the necessity to be vigilant during the entire life cycle of the system while guaranteeing legal compliance, adherence to ethical principles and values, and robustness from both technical and social perspectives (Pos. 30 ff.). Furthermore, ethical principles such as respect for human autonomy, fairness, and explainability are relevant (APA's G*uideline for Artificial Intelligence*, Pos. 72 ff.).

With regards to the *legitimization for the use of AI*, the analyzed documents address the substantial volume of data that falls within the purview of various public communication actors – a task for which automation appears as a resource-saving solution. This encompasses, among other things, the provision of tools designed for the monitoring of self-regulation within the PR and advertising industry or for optimizing the use of resources for journalists. However, "De Volkskrant views artificial intelligence (AI) as a tool, never as a system that can replace the work of a journalist." (*De Volkskrant Protocol*, Pos. 34).

It is necessary to differentiate diverse domains of application in order to attain a comprehensive understanding of the significance of varied mechanisms of responsibility transfer in a hybrid media ecosystem that is progressively reliant on AI systems to (autonomously) generate information and communicate with individuals (Jungherr and Schroeder, 2023). Our sample provides valuable insights into the role of AI in moderating social media platforms, particularly those operated by Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs), as designated by the European Commission in the context of the Digital Services Act (DSA).

> "In order to maintain a safe environment and empower free expression, we remove accounts that are harmful to the community, including those that compromise the security of other accounts and our services. We have built a combination of automated and manual systems to block and remove accounts that are used to persistently or egregiously abuse our Community Standards." (Facebook's *Community Standards*, Pos. 864)

More than half of these community guidelines emphasize transparency as a key principle, and codes often explicitly mention the necessity of human oversight – a feature that is also prioritized by other actors in public communication. Additionally, our analysis identifies a multitude of facets pertaining to prohibited AI application and the corresponding sanctions – encompassing automated mass registration or follow-up procedures as well as artificial influencing of conversations.

The various challenges associated with the *deployment of AI* are not entirely novel: while AI technologies have the potential to facilitate more participatory, pluralistic and dialogic processes through the utilization of personalized bots or news recommenders, they can also be employed to evoke disruptive patterns of communication that have the capacity to foster

polarization and increased legitimacy problems in democracies (Bennett and Livingston, 2018). On the one hand, there appears to be a potential risk of lacking differentiation between synthetic and authentic content, which has the potential to mislead and is exacerbated by the use of artificial manipulation by phenomena such as trolls and malicious bots, deepfakes, and astroturfing. On the other hand, the processes of automation and acceleration are undoubtedly a significant driving force for the creation and dissemination of harmful content, hate speech, and harassment – frequently mentioned in the codes and guidelines. The sample also points to further inductive subcategories, including creativity, content enrichment and experimentation, safeguarding and promoting diversity, and human rights (e.g., non-discrimination) alongside openness and accessibility (of AI systems).

> "Pay particular attention to situations involving more vulnerable groups, such as children, people with disabilities and others who are historically disadvantaged or at risk of exclusion, and to situations characterized by inequalities in power or disposition of information" (Media Perspectives' *Declaration of Intent for Responsible Use of Artificial Intelligence in the Media*, Pos. 13).

Table 2 below presents the key results of our in-depth analysis comparing the documents against the seven principles of trustworthy AI defined in the European Commission's *Ethics Guidelines for Trustworthy Artificial Intelligence*. These guidelines hold a unique position within the sample, primarily due to their clearly defined principles and institutional authority, which provide a solid foundation for thematic analysis. Notably, we identified these principles during our inductive analysis and decided to adopt them as codes. As a result, the guidelines are part of the sample (displayed as a green dot spanning all categories) and served as an analytical lens through which the corpus was systematically structured. The first two rows of table 2 illustrate our coding structure: the type of code represents the main category (see also table 1), while frames of accountability (public, professional, political, and market; Bardoel & d'Haenens, 2004) serve as subcategories, included only when at least one document combines that type of code with the respective frame. Documents may be listed (indicated by a dot) in multiple rows if they are associated with several AI-related themes. However, each document is consistently displayed within a single column corresponding to its designated type-frame-criteria. For simplicity, we differentiate only between national and international documents, without further distinction between individual countries.

Table 2: Principles for trustworthy AI in the main type of codes

	Key requirements (European Commission, 2019)	Policy stakeholders		Journalism				Advertising & PR		Public institutions			Corporate		User	Other
		Political	Public	Market	Professional	Public	Pol.	Market	Pro.	Pro.	Public	Pol.	Market	Public	Market	Market
1	Human agency and oversight	●●●	●	○○○○○○○○○○	○○○●	○		○	●●		○	●		○	○●●●●●	
2	Technical robustness and safety	●●		○	○											
3	Privacy and data governance	●●●	●	○○○○○○○○	○○○○●			○		○	○	●				
4	Transparency	●●●	●●	○○○○○○○○○○○	○○○●●			○			○○				○●●●	●
5	Diversity	●	●	○○○	○●											
	Non-discrimination	●●	●	○○○	○○●								○			
	Fairness	●		○○○○	○●											
6	Societal and environmental well-being	●		○○	○											
7	Accountability via assessment	●●		○○	○●	○					○	●				●
	Accountability via evaluation	●●	●	○○○○○○○○○	○○○●			○		○	○	●				
Total cases per type-frame-combination		4	2	15	6	1	1	1	4	1	6	3	1	2	15	1

○ = document at the national level, ● = document at the international level (source: own data, based on the categories outlined in the European Commission's *Ethics Guidelines for Trustworthy Artificial Intelligence* ●)

Reading example (column-wise): Regarding the professional level of journalism, of six documents in our sample (total cases per type-frame-combination), there were three documents at the national level that dealt with 'human agency and oversight' as well as one international document. The sample contains one example of a journalistic code relevant to the political frame of accountability (the *Dutch Media Act 2008*), which only hints at automated/algorithmic procedures but does not include any reference to the seven categories.

Reading example (row-wise): In addition to the *Ethics Guidelines for Trustworthy Artificial Intelligence* (indicated by the green dot), another 31 documents across eleven out of 15 type-frame-combinations contain text passages referencing the principle of 'human agency and oversight'.

Identifying the principles mentioned in the analyzed documents is relevant for understanding what kind of challenges are acknowledged: when a document refers to a specific principle, it can be interpreted as a recognition of a corresponding challenge. Row-wise, the principles of human agency and oversight (1), transparency (4), and privacy and data security (3) are the principles most frequently stated within the sample. Although the aspects of diversity, non-discrimination and fairness (5) do not seem to be of essential importance in the direct context of AI, it cannot be ruled out that general compliance with these principles was agreed elsewhere in the document. The advantages of AI systems for both citizens and society as a whole are only mentioned in a limited number of analyzed documents, a pattern consistent across all categories within the public frame of accountability. However, the Finnish public broadcaster Yle highlights in its *Principles for Responsible AI* that the development of AI systems is undertaken with the intention of benefiting society, with environmental considerations being of paramount concern (6). Apart from that, societal and environmental well-being as a challenge is only addressed in a few other codes and guidelines. As the ethical considerations of owners and major AI model developers are hardly represented in this sample, statements on technical robustness and safety (2) are rarely found. While some originators advocate for (impact) assessments (7), the most prevalent accountability measure tends to be an ongoing evaluation of those AI systems used (e.g., testing of external service providers, corrections and evaluation of AI sources). In summary, the analyzed documents primarily frame the challenges of AI in public communication through concerns related to human oversight, transparency, and privacy, while broader societal or environmental dimensions are addressed less frequently. This suggests that while immediate operational risks are recognized, more systemic or long-term challenges remain underexplored in many of the reviewed documents.

Emerging governance and accountability mechanisms

Approximately one-third of the analyzed sample consists of professional or organizational journalistic codes (n=23) that focus on the integration of AI as an auxiliary tool for newsroom work processes. Our examination shows that several newsrooms across different countries are increasingly adapting their practices in response to the transformative impact of AI on the industry. The prevailing journalistic codes underscore the significance of upholding established professional standards and responsibilities such as transparency, authenticity, and accuracy, to fulfil media accountability demands (see also Eberwein et al., 2019). These results align with Simon's (2024a) assertion that AI technologies are transforming journalistic practices and optimizing the processes of news production.

The fair use of AI in the journalistic profession is further delineated by the international *Paris Charter on AI and Journalism* by Reporters Without Borders (RSF), which encourages impact assessments and human autonomy, responsibility and oversight, while calling "for AI systems to be designed and used in such a way as to guarantee high-quality, trustworthy, and pluralistic information" (Pos. 38). Notwithstanding the fact that mitigation measures constitute merely a fraction of the gathered codes and guidelines, it is easy to recognize the significance of testing and monitoring mechanisms in providing the operational tools necessary to safeguard these values and address potential vulnerabilities of AI proactively.

While the integration of AI in PR and advertising is generally viewed in a positive light, the documents analyzed primarily emphasize a technocratic approach through investments in automation and AI technologies as means to regulate these fields effectively. The role of AI in content production remains unaddressed, consequently constraining the depth and insight offered by the recommendations, which fail to delve deeply into the intricacies of industry-specific public communication. Although some examples from corporate actors and public institutions were included in the sample due to a reference to AI, their codes and guidelines hardly go into detail regarding the application of specific accountability mechanisms.

Additionally, a closer look at the role of VLOPs and VLOSEs is instructive. On the one hand, VLOPs are forced to implement appropriate moderation procedures that facilitate the identification and removal of content that is deemed inappropriate, harmful or in breach of their community guidelines. Automated procedures seem therefore to be a suitable tool to cope with the enormous amounts of data they must handle. On the other hand, media users are confronted with restrictions on AI content and associated sanctions, including measures such as the blocking or banning of all automation or automated mass logins, and manipulation of conversations (see also Gorwa et al., 2020).

> *"While ethics codes are important, building an ethical culture through public debate, education, and practical learning is essential. Trustworthy AI requires continuous ethical reasoning and sensitivity to contextual details that cannot be captured in general guidelines."* (European Commission's *Ethics Guidelines for Trustworthy AI*, Pos. 102)

The findings of our study underline the European Commission's postulation that integrating AI ethics in public communication necessitates context- and sector-specific adaptations to address complexities that general guidelines and more horizontal approaches are unable to fully capture.

Different approaches by national and supranational stakeholders

While most supranational documents in our sample discuss key principles of responsible AI use from a normative point of view, the everyday application of these principles is mostly relegated to stakeholders at the country level. Additionally, the actors mentioned are less often related to the news media rather than the technology-related sector. An analysis of emerging governance and accountability mechanism therefore needs to consider the national documents as well – and put them in relation to the supranational ones.

A comparison of the documents issued by national and supranational stakeholders shows that they strongly vary in terms of their scope and focus. Codes and guidelines by national organizations are mostly customized to align with the national laws and regulations, different media systems, as well as the cultural, economic and political context. For instance, the Austrian Press Agency APA is subject to the security measures of TÜV AUSTRIA to ensure compliance with quality standards for IT security. Conversely, supranational stakeholders (e.g., the European Commission's *Ethics Guidelines for Trustworthy AI*) prioritize a more extensive framework grounded in fundamental rights, seeking to harmonize shared values and standards across member states from a human-centric approach. By stipulating universal recommendations such as privacy, transparency, accountability, and respect for human rights, these standards can be implemented across the heterogeneous member states. In our sample, these principles are reflected in greater detail in the documents subsumed in the category "policy stakeholders" as summarized in Table 2. These documents mainly offer examples of soft law, co-regulation and policy documents with supranational scope.

Conversely, a marked disparity emerges in the extent to which national-level public communication actors have incorporated AI into their respective codes of conduct and professional guidelines. While certain actors can be identified as pioneers in this domain, the majority of available documents limit themselves to acknowledging the technological and ethical concerns raised by AI, without providing substantive or detailed engagement with these issues. This divergence highlights a broader misalignment between supranational frameworks, typically oriented toward articulating overarching ethical values and principles for the responsible use of AI, and their national-level interpretations, which should ideally translate such abstract frameworks into operational guidelines and follow-up actions regulating the everyday implementation of AI-driven technologies.

Discussion and conclusion

The integration of AI into public communication has introduced both transformative opportunities and complex ethical challenges, as evidenced by our empirical findings. Despite the increasing relevance of AI technologies, their responsible use remains marginally addressed in many codes of ethics across European countries. This discussion focuses on three key aspects: the challenges surrounding ethical codes for AI, the principles and values emphasized in existing guidelines, and the state of governance instruments, with a particular focus on current national and international policy developments.

Challenges of ethical guidelines for AI use

A major challenge in ethical codes related to AI use in public communication lies in their fragmented and often superficial integration of AI-related issues. This result is however not surprising given the still developing and debated guidance both on European and global level. Of the 435 analyzed documents, only a small fraction thoroughly addresses AI and its implications. Many codes fail to consider AI as an all-encompassing issue in self-regulation, reflecting a limited understanding of AI's potential risks and ethical dilemmas – some do not even consider that AI already plays a significant role in their respective field. For instance, countries like Estonia and Slovenia lack any AI-specific codes, while others, such as Austria and the Netherlands, have produced pioneering but isolated documents (e.g., APA's *Guideline for Artificial Intelligence* and Media Perspectives' *Declaration of Intent*). Then again, while the news media industry started a critical self-reflection on the design and use of AI (ANONYMIZED), the advertising industry or public administration have yet to initiate a structural debate about principles for trustworthy AI and about the responsible design and use of AI applications. This uneven landscape underscores the need for broader awareness, discussion, responsible commitment, and integration of AI ethics in media accountability frameworks.

The findings reveal that shared norms and standards for AI governance remain underdeveloped and far from universally established, even within specific industries. In some cases, there is considerable debate about whether AI governance is necessary at all, let alone agreement on a universally accepted framework. This is also reflected in our findings, with codes and guidelines by national stakeholders mostly customized to align with national laws and regulations as well as the cultural, economic and political context of the specific media system. A similar dynamic is observed at the level of supranational institutions. As (ANONYMIZED, forthcoming) demonstrates, ethical guidelines such as the Council of Europe's Guidelines on the Responsible Implementation of Artificial Intelligence Systems in Journalism, the OSCE's Freedom of the Media and Artificial Intelligence report, and UNESCO's Recommendation on the Ethics of Artificial Intelligence all provide a media- specific ethical framework. Yet, while these documents offer valuable insights and recommendations, a universally accepted framework for AI governance in journalism has yet to emerge. This does

not mean that there are no similarities at all. There might be industry-related exceptions: In the news industry, for instance, empirical evidence suggests a certain degree of isomorphism in how AI is regulated. But this happens mostly at the organizational level through codes of ethics, where it is easier to adapt to previously existing guidelines or to those of the market leaders (Becker et al., 2025). Against this backdrop, it remains essential to engage in discussions about the desired principles and values that should underpin AI governance and on how accountability is distributed within the system and among the various actors involved in public communication.

Although the need for international cooperation and shared ethical standards is widely recognized, none of these frameworks provide a clear and actionable roadmap for achieving this goal. Broad regulatory frameworks also face challenges in gaining traction across regions with differing regulatory environments and cultural norms. Moreover, guidelines that are not sector-specific often encounter difficulties in implementation (*ANONYMIZED*, 2023), particularly in the realm of public communication. In this context, the implications for the public sphere are significant, as issues of freedom of expression and press freedom come into play. As Helberger et al. (2020) note, inadequate governance can have potentially detrimental effects on pluralism, privacy, autonomy, and equitable opportunities to communicate.

Principles and values in existing guidelines

The ethical principles highlighted in existing codes/guidelines often include transparency, accountability, fairness, privacy, as well as human agency and oversight. These values are consistently emphasized in supranational frameworks, such as the European Commission's *Ethics Guidelines for Trustworthy AI*, which outline key principles for ensuring human agency, technical robustness, and societal well-being. Our findings echo these priorities, with transparency, human agency and oversight, as well as privacy and data security emerging as the most frequently cited challenges. The findings are also in line with other supranational guidelines produced by the Council of Europe, OCSE, or UNESCO, where human rights, transparency, accountability and ethical considerations are most often mentioned and discussed, but "they exhibit shortcomings, either in terms of applicability (CoE's avoidance of design aspects), or in terms of specificity, actionable guidance and enforceable mechanisms" (Porlezza, 2025: 16). As such here are differences both in the specificity and scope of the specific principles called upon, and therefore also in their recommendations for a governance framework for AI. Empirical evidence from the field of journalism shows that most guidelines "lack comprehensive details on how to address specific ethical principles. Only a limited number of guidelines offer basic and straightforward examples" (de-Lima-Santos et al., 2024). But even specific "ethical cook-book recommendations", based for instance on cases, are risky due to their limited applicability (Porlezza, 2024; Porlezza and Schapals, 2024).

This further demonstrates that calls for more transparency when it comes to the use of AI technology in content creation, for instance through labels, are too simplistic and do not equal

meaningful accountability measures. Instead, transparency is a complex concept that often eludes a concrete translation into everyday practice. However, as Ananny and Crawford (2018: 984) have shown, "[…] transparency alone cannot create accountable systems and engaging with the reasons behind this limitation, we may be able to use the limits of transparency as conceptual tools for understanding how algorithmic assemblages might be held accountable." The limits in the ethical guidelines could therefore also be seen as a starting point for a renewed debate on how the use of AI in public communication should be made transparent and through which instruments algorithmic accountability can be achieved. Nevertheless, transparency is often a slippery slope given that more transparency not necessarily entails more trust, but on the contrary "that audiences perceive news labeled as AI-generated as less trustworthy […]" (Toff and Simon, 2024: 1).

On the other hand, our results also demonstrate that aspects such as societal and environmental well-being are often neglected, indicating a need for a more integrated approach to AI governance. Some forward-thinking documents, such as Yle's *Principles for Responsible AI*, emphasize the societal benefits of AI development, particularly in fostering inclusivity and environmental sustainability. Yet, these examples are exceptions rather than the norm. Nevertheless, the current sketchiness of AI guidelines is not necessarily only a consequence of a *lack* of regulation, as Hunt and McKelvey (2019) have stated. In the case of content platforms for instance, "[…] they are largely left to self-regulate their own inputs and code, [based] on inferred user preferences and the unknowable optimizations chosen by firms or development teams" (idem). Greater emphasis is needed on operationalizing these principles in practical, actionable terms with the goal of developing accountable artificial intelligence (AAI) that enhances rather than undermines democratic values.

Towards a coherent governance framework?

Governance frameworks for AI in public communication are still at an early stage. However, first significant strides have been made at supranational, national, and organizational levels given that until 2024 no regulatory frameworks were in place. At the supranational level, the European Union's AI Act represents a landmark policy effort, focusing on harmonizing AI governance through a horizontal (and therefore not sector-specific) approach across member states. While it primarily targets market regulation, its provisions for high-risk AI applications align with media-specific concerns, such as ensuring human agency and oversight in content moderation. However, critiques of the Act emphasize its limited scope in addressing the unique needs of media and journalism (Helberger and Diakopoulos, 2023; *ANONYMIZED*, 2023). Others such as the Council of Europe, OCSE or UNESCO have developed media-specific ethical guidelines, but they lack a homogeneous governance framework, emphasizing different aspects (*ANONYMIZED*, forthcoming).

This is not to suggest that common principles or mechanisms cannot be identified across the examined documents. Recurrent challenges such as opacity, loss prevention, and privacy

concerns, are frequently addressed through references to transparency, human agency and control, as well as data governance. These recurring elements indicate the presence of dominant principles that shape the discourse. However, a universally accepted framework has not yet materialized, and the field remains in a state of development. Moreover, some documents concentrate specifically on the European regulatory context, which may restrict their broader applicability in regions characterized by different legal and cultural environments. Others adopt a sectoral perspective, focusing on particular market domains, while still others fail to articulate a comprehensive strategy for confronting the global challenges that AI poses in the realm of public communication (Porlezza, 2025).

Despite some advancements, significant gaps remain. Especially in Europe, national- and supranational-level governance instruments frequently lack guidance or consistency, and many fail to address the systemic challenges posed by powerful digital intermediaries like Very Large Online Platforms (VLOPs). These platforms wield significant opinion power: most people nowadays get in touch with news and information via platforms, but the content displayed and recommended on them is selected by algorithms, which are opaque and driven by commercial logic. Switzerland is a good example: as a non-member of the European Union, the Swiss government decided in 2023 that it wants to "strengthen the rights of users in Switzerland and demand more transparency from the platforms without limiting their positive influence on freedom of expression" (Federal Council, 2023). But once again, the main question will be whether and to what extent the Federal Council will opt for a coherent and integrated governance framework by the European Union's Digital Services Act. Even more so, as the Swiss government decided at the beginning of 2025 to delay the deliberation on platform regulation following Trump's second term as president of the United States, only to launch the consultation process on October 29, 2025 (AlgorithmWatch, 2025). Other European countries instead, such as Slovenia, have now developed a national governance through adopted laws that regulate the use of (generative) AI (ANONYMIZED).

Overall, a more integrated and dynamic governance framework is needed. This framework should combine regulatory oversight with self-regulation and emphasize multi-stakeholder engagement to achieve accountable artificial intelligence in public communication. In conclusion, while progress has been made in articulating the principles and values underpinning responsible AI use (see for instance the Steering Committee on Media and Information Society's (CDMSI) Guidelines on the Responsible Implementation of Artificial Intelligence Systems in Journalism; or the Swiss Press Council's guidelines on the use of AI), the implementation of these principles and their embedding into accountability measures remain inconsistent and incomplete. To ensure that AI contributes positively to public communication, future efforts must focus on bridging the gap between high-level ethical aspirations and practical as well as operational governance frameworks. This requires ongoing collaboration among policymakers, media organizations, technologists, and civil society to create a balanced and inclusive AI governance landscape that safeguards democracy and upholds the fundamental values of public communication.

Like all empirical research, this study has limitations. For example, the utilization of automated translation software during our analysis of documents from various national backgrounds may have resulted in the occurrence of inaccuracies, particularly in the context of minor national languages. However, considering that a substantial proportion of the sample was originally published in English, the impact of potential linguistic mistakes seems not very substantial.

In contrast to the considerable number of public communication codes and guidelines identified in the *ANONYMIZED* project, only a limited number of documents contained pertinent passages regarding AI – a smaller proportion of these documents constituted the primary focus of our analysis. Consequently, the *ANONYMIZED PROJECT NAME* sample merely alludes to the perspectives of other actors in the domain of public communication, such as PR, advertising and media users. To facilitate a more profound comprehension of a specific actor, it would be helpful if future scholarly endeavors extended this analysis to encompass, for instance, national community and netiquette guidelines.

Building on this, future research could also focus on how international cooperation can evolve to establish internationally shared and recognized standards that ensure the responsible use of AI applications in public communication. Such cooperation is essential to mitigate dysfunctional effects on citizens, the public sphere, and democracy. Moreover, the ethical considerations surrounding AI often overlook the critical stages of design and development, even though these stages significantly shape the technology's societal impact. This includes examining how design choices can address minorities, diverse cultural contexts and perspectives to ensure that AI technologies enhance democratic values, protect individual rights such as freedom of expression, and foster inclusive public communication.

[1] https://*anonymized* (see also anonymized, 2024).
[2] The following search terms were included: AI* (AI-powered; AI-driven; generative AI); algorithm*; artificial* (intelligence); astroturfing; automated*; automation* (intelligent automation; robotic process automation); computational*; data* (big data, data mining); emerging technology*; human*-first strategy; machine learning*; (un-)supervised learning; deep* learning; natural language processing (NLP); smart technologies.
[3] Documents published in languages other than English were translated with the help of DeepL to enable a collaborative analysis within the multi-national *ANONYMIZED PROJECT NAME* research team.
[4] The final version of the sample for this research project was compiled on September 30, 2024. Any documents updated after that date could not be taken into account. Hence, the *ICC Advertising and Marketing Communications Code* was analyzed based on its version from 2018, which briefly considered the topic of AI, while the more elaborate 2024 version was not part of this study.

Funding

This study has received funding from the European Union's Horizon Europe research and innovation program under the grant agreement No *anonymized* (within the project *ANONYMIZED PROJECT NAME*). The work reflects only the authors' views, and the Commission is not responsible for any use that may be made of the information it contains.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Abelmann C and Elmore R (1999) *When accountability knocks will anyone answer?* Philadelphia: Philadelphia Consortium for Policy Research in Education.

Ahmad R (2009) Governance, social accountability, and civil society. Liberalism bypassing politics? In: Singh A, Kapoor K and Bhattacharyya R (eds) *Governance and poverty reduction: Beyond the cage of best practices*. New Delhi: PHI Learning Private Limited, pp. 310–337.

AlgorithmWatch (2025) Regulierung von Online-Plattformen: Was tut die Schweiz? Available at: https://algorithmwatch.ch/de/leitfaden-plattformregulierung-schweiz/

Ananny M and Crawford K (2018) Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20(3): 973–989.

Bardoel J and d'Haenens L (2004) Media responsibility and accountability: New conceptualizations and practices. *Communications* 29: 5–25.

Becker KB, Simon FM and Crum C (2025) Policies in parallel? A comparative study of journalistic AI policies in 52 global news organizations. *Digital Journalism*, 1-21. DOI: 10.1080/21670811.2024.2431519.

Bennett WL and Livingston S (2018) The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication* 33(2): 122–139. DOI: 10.1177/0267323118760317.

Braun V and Clarke V (2012) Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D and Sher KJ (eds) *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological*. Washington DC: American Psychological Association, pp. 57–71. DOI: 10.1037/13620-004.

Chadwick A (2017) *The hybrid media system: Politics and power*. Oxford: Oxford University Press.

Coeckelbergh M (2024) *Why AI undermines democracy and what to do about it*. New Jersey: Wiley.

Corrêa NK, Galvão C, Santos JW, Del Pino C, Pinto EP, Barbosa C, Massmann D, Mambrini R, Galvão L, Terem E and de Oliveira N (2023) Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns* 4(10): 100857–100857. DOI: 10.1016/j.patter.2023.100857.

Council of Europe (2023) *Guidelines on the responsible implementation of artificial intelligence (AI) systems in journalism*. https://www.coe.int/en/web/freedom-

expression/-/guidelines-on-the-responsible-implementation-of-artificial-intelligence-ai-systems-in-journalism

Dafoe A (2024) AI governance. In: Bullock JB, Chen Y-C, Himmelreich J, Hudson VM, Korinek A, Young MM and Zhang B (eds) *The Oxford handbook of AI governance*. Oxford: Oxford University Press, p. 21.

de-Lima-Santos MF and Ceron W (2021) Artificial intelligence in news media: Current perceptions and future outlook. *Journalism and Media* 3(1): 13–26.

de-Lima-Santos MF, Yeung WN and Dodds T (2024) Guiding the way: A comprehensive examination of AI guidelines in global media. *AI & Society:* 1–19. DOI: 10.1007/s00146-024-01973-5.

Diakopoulos N (2019) *Automating the news: How algorithms are rewriting the media*. Cambridge, MA: Harvard University Press.

Diakopoulos N and Koliska M (2017) Algorithmic transparency in the news media. *Digital Journalism* 5(7): 809–828.

Dobek-Ostrowska B, Głowacki M, Jakubowicz K and Sükösd M (eds) (2010) *Comparative media systems: European and global perspectives*. Budapest, New York: Central European University Press.

Eberwein T, Fengler S and Karmasin M (eds) (2019) *Media accountability in the era of post-truth politics: European challenges and perspectives*. London, New York: Routledge.

Edwards M and Hulme D (1996) *Beyond the magic bullet: NGO performance and accountability in the post-Cold War world*. London: Earthscan.

European Commission (2019) *Ethics Guidelines for Trustworthy Artificial Intelligence.* Report, High-Level Expert Group on AI, 8 April 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Parliament and Council (2024) *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) 2018/858, (EU) 2019/2144 and (EU) 2019/881 and Directives 2014/90/EU, 2014/53/EU, 2006/42/EC and (EU) 2020/1828* (Text with EEA relevance). Official Journal of the European Union, L 202, 1–164. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Fengler S, Eberwein T, Alsius S, Baisnée O, Bichler K, Dobek-Ostrowska B, Evers H, Glowacki M, Groenhart H, Harro-Loit H, Heikkilä H, Jempson M, Karmasin M, Lauk E, Lönnendonker J, Mauri M, Mazzoleni G, Pies J, Porlezza C, … and Zambrano SV (2015) How effective is media self-regulation? Results from a comparative survey of European journalists. *European Journal of Communication 30*(3): 249–266. DOI: 10.1177/0267323114561009.

Fengler S, Eberwein T, Karmasin M, Barthel S and Speck D (2022) Media accountability: A global perspective. In: Fengler S, Eberwein, T and Karmasin M (eds) *The global handbook of media accountability*. London, New York: Routledge, pp. 3–57. DOI: 10.4324/9780429326943-2.

Gorwa R, Binns R and Katzenbach C (2020) Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1). DOI: 10.1177/2053951719897945.

Gunkel, D. (2020). *An introduction to communication and artificial intelligence.* Cambridge: Polity Press.

Guzman AL and Lewis SC (2020) Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society* 22(1): 70–86. DOI: 10.1177/1461444819858691.

Guzman AL and Lewis SC (2024) What generative AI means for the media industries, and why it matters to study the collective consequences for advertising, journalism, and public relations. *Emerging Media* 2(3): 347–355. DOI: 10.1177/27523543241289239.

Hallin DC and Mancini P (eds) (2012) *Comparing media systems beyond the Western world*. New York: Cambridge University Press.

Hargreaves A (2008) The fourth way of change: Towards an age of inspiration and sustainability. In: Hargreaves A and Fullan M (eds) *Change wars.* Toronto: Solution Tree, pp. 11–44.

Helberger N, Eskens S, van Drunen M, Bastian M and Moeller J (2019) *Implications of AI-driven tools in the media for freedom of expression.* Report by the Institute for Information Law (IViR). Available at: https://rm.coe.int/coe-ai-report-final/168094ce8f (accessed 15.01.2025).

Helberger N, van Drunen M, Eskens S, Bastian M and Moeller J (2020) A freedom of expression perspective on AI in the media – with a special focus on editorial decision making on social media platforms and in the news media. *European Journal of Law and Technology* 11(3). DOI: 10.31235/osf.io/c4af9.

Hunt R and McKelvey F (2019) Algorithmic regulation in media and cultural policy: A framework to evaluate barriers to accountability. *Journal of Information Policy* 9: 307–335.

Jobin A, Ienca M and Vayena E (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence 1*(9): 389–399.

Jungherr A (2023) Artificial intelligence and democracy: A conceptual framework. *Social Media + Society* 9(3). DOI: 10.1177/20563051231186353.

Jungherr A and Schroeder R (2023) Artificial intelligence and the public arena. *Communication Theory* 33(2–3): 164–173.

Kaldor M (2003) Civil society and accountability. *Journal of Human Development* 4(1): 5–27. DOI: 10.1080/1464988032000051469.

Kaplan J (2025) More speech and fewer mistakes. Available at: https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/ (accessed 10 January 2025).

Kõuts-Klemm R, Eberwein T, Peruško Z, Vozab D, Rožukalne A, Skulte I and Stakle A (2024) Media and journalism research in small European countries. *Media and Communication 12*, Article 7205. DOI: 10.17645/mac.7205

Landman T (2008) *Issues and methods in comparative politics*. London: Routledge.
Mitova E, Blassnig S, Strikovic E, Urman A, Hannak A, de Vreese CH and Esser F (2023) News recommender systems: A programmatic research review. *Annals of the International Communication Association* 47(1): 84–113.
Natale S (2021) *Deceitful media: Artificial intelligence and social life after the Turing test*. Oxford: Oxford University Press.
Novelli C, Taddeo M and Floridi L (2024) Accountability in artificial intelligence: What it is and how it works. *AI & Society* 39(4): 1871–1882.
Organization for Economic Co-operation and Development (2019) *OECD principles on artificial intelligence*. https://www.oecd.org/going-digital/ai/principles/
Painter M (2007). Defining accountability in a network society. *Business Ethics Quarterly* 17(3): 515–534. DOI: 10.2307/27673194.
Peruško Z, Vozab D and Čuvalo A (2020) *Comparing post-socialist media systems: The case of Southeast Europe*. Abingdon, New York: Routledge.
Porlezza C (2023) Promoting responsible AI: A European perspective on the governance of artificial intelligence in media and journalism. *Communications 48*(3): 370–394.
Porlezza C (2025) Artificial intelligence's implications for freedom of expression: Different institutional approaches to ensure a responsible use of AI in journalism. *Ethical Space: International Journal of Communication Ethics 22*(1-2): 1–19. doi:10.21428/0af3f4c0.031568dc.
Porlezza C and Schapals AK (2024) AI ethics in journalism (studies): An evolving field between research and practice. *Emerging Media 2*(3):356–370.
Prior L (2003) *Using documents in social research*. London: Sage.
Rozgonyi K (2023) Accountability and platforms' governance: The case of online prominence of public service media content. *Internet Policy Review*, *12*(4). DOI: 10.14763/2023.4.1723.
Sahlberg P (2010) Rethinking accountability in a knowledge society. *Journal of Educational Change* 11: 45–61. DOI: 10.1007/s10833-008-9098-2.
Sahlberg P (2007) Education policies for raising student learning: The Finnish approach. *Journal of Education Policy* 22(2): 147–171. DOI: 10.1080/02680930601158919.
Schultz MD and Seele P (2023) Towards AI ethics' institutionalization: Knowledge bridges from business ethics to advance organizational AI ethics. *AI and Ethics* 3(1): 99–111.
Simon FM (2022) Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism* 10(10): 1832–1854.
Simon F (2024a) *Artificial intelligence in the news: How AI retools, rationalizes, and reshapes journalism and the public arena*. New York: Columbia University/Tow Center.
Simon FM (2024b) Escape me if you can: How AI reshapes news organizations' dependency on platform companies. *Digital Journalism* 12(2): 149–170.
Steinhoff J (2023) AI ethics as subordinated innovation network. *AI & Society* 39: 1995–2007. DOI: 10.1007/s00146-023-01658-5.

Sutrop M (2020) Challenges of Aligning Artificial Intelligence with Human Values. *Acta Baltica Historiae et Philosophiae Scientiarum 8*(2): 54–72.

Toff B and Simon FM (2024) "Or they could just not use it?": The dilemma of AI disclosure for audience trust in news. *The International Journal of Press/Politics*, DOI: 10.1177/19401612241308697.

Ulnicane I (2025) Governance fix? Power and politics in controversies about governing generative AI. *Policy and Society*, *44*(1): 70–84. DOI: 10.1093/polsoc/puae022

United Nations Educational, Scientific and Cultural Organization. (2021). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000380455

Van Es K and Nguyen D (2024) Balancing needs and values: A multi-stakeholder examination of algorithmic news recommenders in the Netherlands. *Journalism Practice*: 1–19. DOI: 10.1080/17512786.2024.2392654.

VERBI Software (2021) MAXQDA 2022 [computer program]. Berlin: VERBI Software.

Verdegem P (2021) *AI for everyone? Critical perspectives.* London: University of Westminster Press.

Wagner, B (2018) Ethics as an escape from regulation. From "ethics-washing" to ethics-shopping? In: Bayamlioğlu E, Baraliuc, I, Janssens L et al. (eds) *Being Profiled: Cogitas Ergo Sum. 10 Years of 'Profiling the European Citizen'.* Amsterdam: Amsterdam University Press, pp. 84–88. DOI: https://doi.org/10.25969/mediarep/13281.

Vogler D, Eisenegger M, Fürst S, Udris L, Ryffel Q, Rivière M and Schäfer M (2023) Artificial intelligence in news production: Perception and acceptance among the Swiss population. In: fög (eds) *Yearbook quality of the media 2023*. Basel: Schwabe.

Vogler, D., Fürst, S., Ryffel, Q., Eisenegger, M., Porlezza, C., & Procopio, N. (2024). Governance von KI im Journalismus. Report for the Federal Office of Communication (OFCOM).

Zednik C (2021) Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. *Philosophy and Technology 34*(2): 265–288. https://doi.org/10.1007/S13347-019-00382-7